RT-PCR-BASED CLONING OF HUMAN SMN, THE SMA DETERMINING GENE, AND THE CONSTRUCTION OF ITS EXPRESSION PLASMIDS

INVENTOR:

KHUE VU NGUYEN, Ph.D.

2828 University Avenue, Apt # 303

San Diego, California 92104

U.S.A.

Tel.: (619) 299-0449

Fax: (619) 543-7868

Email: kv52nguyen@yahoo.com

CORRESPONDENCE:

Dr. HUYNH-MAI T. NGUYEN

7222 Via Bella

San Jose, CA 95319

U.S.A.

Tel: (408) 281-4378; (408) 482-0519

Fax: (408) 924-3775

Email: bulan_99@yahoo.com

RT-PCR-BASED CLONING OF HUMAN SMN, THE SMA DETERMINING GENE, AND THE CONSTRUCTION OF ITS EXPRESSION PLASMIDS

I. FIELD OF THE INVENTION

Spinal muscular atrophy (SMA) is a lethal autosomal recessive disease. SMA is characterized by the degeneration of motor neurons in the spinal cord, causing progressive weakness of the limbs and trunk, followed by muscle atrophy. The gene most highly associated with SMA is the survival motor neuron (SMN) gene. The availability of full-length human SMN protein will be valuable for biochemical and biological analyses that may elucidate the molecular mechanism of SMA. The present invention is a new approach to the cloning of human SMN gene based on the reverse transcription (RT) and the polymerase chain reaction (PCR), and the construction of expression plasmids in order to obtain full-length human SMN protein.

II. BACKGROUND OF THE INVENTION

Spinal muscular atrophy (SMA) is characterized by the degeneration of anterior horn cells of the spinal cord, leading to progressive symmetrical limb and trunk paralysis and muscular atrophy. It is the second most common fatal autosomal recessive disorder after cystic fibrosis and is the common genetic cause of childhood mortality (1-4). Childhood spinal muscular atrophies are divided into three clinical groups on the basis of age of onset and clinical course (5). The so-called Werdnig-Hoffman disease (6,7), or SMA type I (SMA-I), is characterized by severe, generalized muscle weakness and hypotonia that is noticeable at birth or appears within the first 3 months of life. In this acute form of the disease, death by respiratory failure usually occurs within the first 2 years. Type I SMA patients are never able to sit or walk.

Children who have the intermediate form, SMA type II (SMA-II), can sit but cannot stand or walk unaided. They survive beyond 4 years of age until adolescence or later. Patients with Kugelberg-Welander disease (8), or SMA type III (SMA-III), display proximal muscle weakness that begins to develop after 2 years of age and able to stand and walk, but often become wheelchair-bound during youth or adulthood.

All three types of SMA map to chromosome 5q 13.3, suggesting that they are the result of allelic mutations (9-12). The genomic region encompassing the disease gene is particularly unstable and prone to large scale deletions (13). The molecular mechanism of motor neuron death in SMA remains unknown, but two genes associated with SMA have been identified in the following region: Survival motor neuron, SMN (14) and neuronal apoptosis inhibitory protein (NAIP) (15). Lefebvre et al. (14) identified the SMN gene with 8 exons extending over approximately 20 kb. It was found that the SMN gene was either absent or interrupted in its exons 7 and 8 in the majority of patients (98%), independently of the type of SMA, suggesting that SMN is the determining gene (14). Roy et al. (15) identified the NAIP gene with 16 exons spanning 60 kb. It was shown that 45% of SMA type I and 18% of SMA type II and III patients have a partial or complete homozygous deletions of the NAIP gene. Since the NAIP gene is very close to the SMN gene, it is hypothesized that the mutations in the SMN gene are the major determinant of the SMA phenotype, whereas the extent of the deletions, which may include the NAIP gene, may modify the SMA phenotype thus accounting for the different clinical subtypes of the disease (14, 15). The SMN gene encodes a 38-kDa protein ubiquitously expressed, the SMN protein, which reveals no significant sequence homology with other protein. Its exact function is actually unknown, but it interacts with several other proteins involved in spliceosomal assembly, suggesting a role for SMN in pre-mRNA splicing (16-18). SMN is an essential

protein since a knockout mouse for the SMN gene is lethal embryonically (19). A study of the distribution of the SMN protein complex in human fetal tissues showed interesting results in favor of the "muscle hypothesis" of SMA (20). In fetal muscle cells, SMN protein was concentrated in large cytoplasmic dot-like structures similar to the size of a gem (the nuclear structure containing the SMN protein), which were never previously described in the cytoplasm. This immunolocalization of SMN protein in skeletal muscle contrasts with results obtained in other normal tissues, such as thymus, kidney, lung, and brain or liver and spinal cord (21). Animal models for SMA are recent (22-24). Thus, the availability of full-length SMN protein is valuable for biochemical and biological analyses that may elucidate the molecular mechanism of SMA. Currently, the procedure to obtain full-length SMN protein is the isolation of SMN-cDNA clones from the cDNA library using the synthesized oligonucleotide probes; this method of probing to identify appropriate SMN-cDNA for cloning is costly, very time-consuming (requiring labor intensive steps), and requires highly skilled personnel. Consequently, very few laboratories have the equipment or capability to produce full-length SMN protein; for that reason, other laboratories, which do not have such capability, would have to obtain SMN-cDNA clones from the few laboratories that can produce SMN-cDNA clones. Therefore, there is a need to design an easy, simple and cost-effective procedure to obtain SMN-cDNA for cloning. For such a purpose, in the present invention, a different approach is used: Instead of using probes to isolate SMN-cDNA clones, the new procedure involves starting with the SMN-messenger ribonucleic acid (SMN-mRNA) and utilizing the RT and PCR reactions in order to obtain SMN-cDNA for cloning. This is a simple and cost-effective procedure that can be performed in any laboratory.

In addition, the cloning of SMN-cDNA based on RT-PCR reactions allows the construction of expression plasmids, (a) using the pFastBac™ HTb and the pBlueBacHis2 A transfer vectors for the purpose of obtaining human SMN protein in insect cells; and (b) using the pET-28a (+) transfer vector for the purpose of obtaining human SMN protein in bacteria.

III. PURPOSE OF THE INVENTION

The purpose of the invention is: 1/ To perform the procedure to obtain SMN-cDNA for cloning by starting with SMN-messenger ribonucleic acid (SMN-mRNA) and by utilizing the RT and PCR reactions; and 2/ to construct expression plasmids (a) using the pFastBac™ HTb and the pBlueBacHis2 A transfer vectors for the purpose of obtaining human SMN protein in insect cells; (b) using the pET-28a (+) transfer vector for the purpose of obtaining human SMN protein in bacteria. The availability of full-length SMN protein is essential for the identification of biochemical and biological activities that occur in the SMN protein which will allow the exploration of gene therapy in SMA.

IV. MATERIALS AND METHODS

Isolation of RNA

The total ribonucleic acid (RNA) was isolated from a sample of human liver biopsy (wild type) according to the method described by Sambrook et al. (25) using guanidin/phenol (Tris Reagent™, Euromedex, 67460 Souffelweyersheim, France). The total RNA was dissolved in water pre-treated by 0.1% diethyl pyrocarbonate (DEPC, Sigma, St. Louis, MO). This RNA solution is ready for subsequent treatment for the synthesis of the cDNA.

Reverse Transcription

The synthesis of the cDNA was performed by reverse transcription (RT), as described by Sambrook et al. (25). The first copies of cDNA were synthesized by using the synthesized oligonucleotide (SEQ ID NO. 1) (Genosys Biotechnologies, Europe, Ltd., France) with the following sequence: 5' TGGCAGACTTAC 3' (SEQ ID NO. 1). This oligonucleotide was selected by taking the complementary sequence between base pairs 889 and 900 of the human mRNA for SMN described by Lefebvre et al. (14). The M-MLV Reverse Transcriptase enzyme (Gibco BRL^R, Life Technologies Sarl, BP 96, 95613 Cergy Pontoise, France) was used in order to perform the reverse transcription reaction. This reaction was performed as follow:

0.1 nmol of oligonucleotide SEQ ID NO. 1, and 10 pmol each of nucleotides dATP, dCTP, dGTP and dTTP were added to 5 µg of the total RNA. The reaction was conducted in the presence of a reaction buffer for RT of the Gibco BRL^R kit; the total volume of the reaction was 20 µl. After heating the mixture to 90^{0}C for 2 minutes and then cooling it on ice for 1 minute, 200 U M-MLV were added; then the mixture was left at 25^{0}C for 10 minutes and then at 42^{0}C for 45 minutes.

Amplification

Having obtained the first copies of cDNA by RT reaction, the next step is to amplify the RT products.

Amplifying the RT products were assessed by using the polymerase chain reaction (PCR) technique (26, 27). Two synthesized oligonucleotides (SEQ ID NO. 2) (forward primer) and (SEQ ID NO. 3) (reverse primer) (Genosys Biotechnologies) were used in order to perform the PCR reaction. They have the following sequences:

5' ATGGCGATGAGCAGCGG 3' (SEQ ID NO. 2) and

5' TTAATTTAAGGAATGTGAGCAC 3' (SEQ ID NO. 3)

The oligonucleotide SEQ ID NO. 2 was based on the sequence between base pairs 1 and 17 of the human mRNA sequence for SMN described by Lefebvre et al. (14). The oligonucleotide SEQ ID NO. 3 was selected by taking the complementary sequence between base pairs 864 and 885 of the human mRNA sequence for SMN described by Lefebvre et al. (14). Amplification was conducted by using a DNA Thermal Cycle (AmplitronR II Thermolyse). The reaction was conducted in a total volume of 50 µl with 2.5 U of Taq DNA polymerase (Promega Corporation, Madison, WI, U.S.A.) in the presence of the PCR reaction buffer from Promega kit containing 0.1 nmol of oligonucleotide (SEQ ID NO. 2) and 0.1 nmol of oligonucleotide (SEQ ID NO. 3), 10 pmol each of nucleotides dATP, dCTP, dGTP, and dTTP, 62.5 pmol of $MgCl_2$ (Promega) and 5 µl of the reverse transcription reaction medium obtained previously. Amplification conditions were as follow: Denaturating at 94^0C for 1 minute, anneling at 55^0C for 2 minutes, and elongating at 72^0C for 1 minute, each for 35 cycles. The PCR product was analysed by electrophoresis on a 20 g/l agarose gel to screen for the presence of the appropriate-size band by using the fluorescent dye ethidium bromide. The PCR product was then isolated, purified by phenol-chloroform extraction, dried and resuspended in distilled water according to the method described by Sambrook et al. (25).

Cloning

The obtained purified PCR product was then subjected to the ligation reaction into the pCR^R II plasmid vector of the TA Cloning kit (Invitrogen). The reagents of this kit and the reaction conditions performed according to the manufacturer's recommendations were used. The ligation product was then introduced in INVαF' E. Coli strain by using the reagents and the transformation procedure of the TA Cloning kit (Invitrogen). The screening for inserts was

performed by using blue-white color selection. The sequencing of inserts obtained was performed by using the ABI DNA sequencer. The resulting vector was termed (1) (pCR^R II/SMN-cDNA).

Construction of the Expression Plasmids for Human SMN Protein

To construct the expression plasmids for human SMN protein, the following systems for expression of SMN recombinant protein were performed:

1. Insect Expression Systems

1. 1. Using the Bac-to-Bac^R Baculovirus Expression System

From the vector (**1**) obtained previously, the BamHI to XhoI fragment containing the cDNA coding sequences of SMN protein was isolated and subjected to the ligation reaction into the baculovirus transfer vector pFastBac^{TM} HTb (Invitrogen) predigested by BamHI and XhoI and predephosphorylated with calf intestinal alkaline phosphatase (Boerhinger Mannheim, GmbH, Germany). The reaction was conducted in the presence of the reagents for the ligation of the TA Cloning kit (Invitrogen); the reaction conditions used are according to the manufacturer's recommendations. The ligation product was then introduced in INVαF' E. Coli strain (Invitrogen) by using the reagents and the transformation procedure of the TA Cloning kit (In vitrogen). The screening for inserts was based on the presence of white colonies. The resulting vector thus obtained was termed (**2**) (pFastBac^{TM} HTb/SMN-cDNA). This vector was then introduced in the DH10Bac^{TM} E. Coli competent cells (Invitrogen) by using the reagents and the transformation procedure of the Bac-to-Bac^R Baculovirus Expression System (Invitrogen). The screening for recombinant bacmids was performed by using blue-white color selection. The verification of the presence of SMN-cDNA's insert in the recombinant bacmid was performed by PCR amplification using the M13 forward (-40) and M13 reverse primers (Invitrogen). The

reaction conditions used are according to the manufacturer's recommendations. The PCR product was analysed by electrophoresis on a 20 g/l agarose gel to screen for the presence of the appropriated-size band by using the fluorescent dye ethidium bromide. The PCR product was then isolated, purified by phenol-chloroform extraction, dried and resuspended in distilled water according to the method described by Sambrook et al. (25). The sequencing of SMN-cDNA's insert was performed by using the ABI DNA sequencer. The recombinant bacmid thus obtained was termed (**3**).

1. 2. Using the Bac-N-Bac™ Baculovirus Expression System

From the vector (**2**) obtained previously, the BamHI to XhoI fragment containing the cDNA coding sequence of SMN protein was isolated and subjected to the ligation reaction into the baculovirus transfer vector pBlueBacHis2 A (Invitrogen) predigested by BamHI and XhoI and predephosphorylated with calf intestinal alkaline phosphatase (Boehringer Mannheim, GmbH, Germany). The reaction was conducted in the presence of the reagents for the ligation of the TA Cloning kit (Invitrogen); the reaction conditions used are according to the manufacturer's recommendations. The ligation product was then introduced in INVαF' E. Coli strain (Invitrogen) by using the reagents and the transformation procedure of the TA Cloning kit (Invitrogen). The screening for inserts was performed by using blue-white color selection. The resulting vector thus obtained was termed (**4**) (pBlueBacHis2 A/SMN-cDNA).

2. Prokaryotic Expression Systems

From the vector (**2**) obtained previously, the BamHI to XhoI fragment containing the cDNA coding sequences of SMN protein was isolated and subjected to the ligation reaction into the pET-28a (+) plasmid vector (Novagen) predigested by BamHI and XhoI and predephosphorylated with calf intestinal alkaline phosphatase (Boehringer Mannheim, GmbH,

Germany). The reaction was conducted in the presence of the reagents for the ligation of the TA Cloning kit (Invitrogen); the reaction conditions used are according to the manufacturer's recommendations. The ligation product was then introduced in INVαF' E. Coli strain (Invitrogen) by using the reagents and the transformation procedure of the TA Cloning kit (Invitrogen). The screening for inserts was based on the presence of white colonies. The resulting vector thus obtained was termed (**5**) (pET-28a (+)/SMN-cDNA).

V. RESULTS AND DISCUSSION

Cloning of Human SMN Gene

In order to investigate the molecular mechanism of SMA, intense efforts have been directed toward the identification of SMN-cDNA clones which encode the amino acid sequences of human SMN (14). The availability of the SMN-cDNA clones has permitted the construction of the expression plasmid pEGFP-C3/SMN for human SMN recombinant protein (28). Up to now, the only approach to identify SMN-cDNA clones from the cDNA library is by using the synthesized oligonucleotide probes (14); however, this approach is costly, time-consuming (requiring labor intensive steps) and it requires highly skilled personnel. Furthermore, not many laboratories have the equipment or capability to perform such a procedure. In contrast, the present invention is the development of a procedure that is easy, simple and cost-effective which can be performed in any laboratory; this procedure starts with the human SMN-mRNA and includes the utilization of the RT and PCR reactions in order to obtain SMN-cDNA for cloning. As shown in Figure 1, the PCR product of ~ 0.9 kb of the human SMN gene was successfully amplified by two synthesized oligonucleotides (SEQ ID NO. 2) and (SEQ ID NO. 3). The obtained PCR product was subcloned into the pCR^R II plasmid vector of 3.9 kb (Invitrogen).

The analysis of the sequence of inserts showed that the DNA sequences of the PCR product completely matched with the sequences of SMN-mRNA described by Lefebvre et al. (14). Then, the orientation of SMN-cDNA's insert in the plasmid vector (**1**) (pCR^R II/SMN-cDNA) (Figure 2) was determined by the digestion of (**1**) with NcoI and HindIII; the result shows two fragments of 4,220 and 565 bp (data not shown).

Construction of the Expression Plasmids for Human SMN Protein

A schematic representation of the different cloning steps for constructing the expression plasmids for human SMN recombinant protein are given in Figures 3-5.

1. Insect Expression Systems

Insect expression systems are relative newcomers to the gene expression arena. As higher eukaryotes, insects cells offer many mammalian posttranslational modifications. Insect cells have proven to be an excellent host for recombinant protein expression. They are often chosen for protein production because they are a higher eukaryotic with easy cell culture and can be readily adapted to high-density suspension culture for large-scale expression.

1.1 Using the Bac-to-Bac^R Baculovirus Expression System

The Bac-to-Bac^R Baculovirus Expression system is the fastest route available for producing recombinant baculovirus. This system uses a unique bacmid shuttle vector that combines with an expression cassette by siote-specific transposition to create an expression bacmid. The expression bacmid is then transfected into insect cells to generate high-titer, ready-to-use recombinant baculovirus. With Bac-to-Bac^R, recombinant protein can be expressed in as little as nine days. The Bac-to-Bac^R HT baculovirus expression system (Invitrogen) is used. This system offers the components necessary for expression and purification of histidine tagged recombinant proteins in Sf9, Sf21, or High Five^{TM} cells. The system includes the pFastBac^{TM}

HTb vector of 4.8 kb, the DH10Bac™ E. Coli competent cells for cloning. Expression of the recombinant fusion proteins from the pFastBac™ HTb vector is driven by the polyhedrin promotor. Proteins expressed are fused at the N-terminus to a tag of six tandem histidine residues and a TEV protease cleavage site for removal of histidine tag following protein purification. The histidine residues create a high-affinity metal binding site to allow purification of recombinant fusion proteins on nicken-chelating resin. In the present work, a schematic representation of the different cloning steps for constructing the expression plasmid for SMN protein using the pFastBac™ HTb vector is given in Figure 3. The correct orientation for SMN protein expression of the SMN-cDNA's insert in the plasmid vector (**2**) was confirmed by the digestion of (**2**) with NcoI which gave two fragments of 5,120 and 565 bp (data not shown). The verification of the presence of SMN-cDNA's insert in the recombinant bacmid (**3**) using PCR technique (in the presence of M13 forward (-40) and M13 reverse primers) gave a DNA fragment of 3,315 bp (data not shown). The DNA sequences of the obtained PCR product completely matched with the sequences of SMN-mRNA described by Lefebvre et al. (14).

1.2. Using the Bac-N-Bac™ Baculovirus Expression System

The Bac-N-Bac™ Baculovirus Expression system has been used for over a decade to produce high levels of recombinant proteins. The baculovirus transfer vector pBlueBacHis2 A (Invitrogen) is used. The pBlueBacHis2 A is a 4.9 kb polyhedrin promotor-based vector that is designed to create N-terminal fusion proteins produced in the baculovirus system. The vector's small size allows an easier cloning of the gene of interest. Expression of β-galactosidase following recombination with Bac-N-Blue™ Linear DNA promotes simplified screening of recombinant plaques. Proteins expressed from pBlueBacHis2 A are fused at the N-terminus to a tag of six tandem histidine residue and an enterokinase cleavage site (the Xpress™ tag). The

histidine residues create a high-affinity metal binding site to allow purification of recombinant fusion proteins on nickel-chelating resin. The Xpress™ tag is easily cleaved away from the protein using enterokinase. Expression and purification of the Xpress™ fusion protein is easily tracked using the Anti-Xpress™ Antibody which recognizes an epitope located in the Xpress™ tag. In the present work, a schematic representation of the different cloning steps for constructing the expression plasmid for SMN protein using the pBlueBacHis2 A vector is given in Figure 4. The correct orientation for SMN expression of the SMN-cDNA's insert in the plasmid vector (**4**) was confirmed by the digestion of (**4**) with NcoI which gave two fragments of 5,465 and 320 bp (data not shown).

2. Prokaryotic Expression Systems

Prokaryotic hosts, especially E. Coli, have been widely used for the expression of recombinant proteins. The Prokaryotic Expression systems offers several advantages as recombinant protein expression hosts, including easy manipulation, rapid growth and simple media requirements. Prokaryotic expression systems are well suited for expression of proteins that will be used in antibody production and for structural studies. The Prokaryotic Expression systems use the T7 expression system which allows high-level expression from the strong bacteriophage T7 promoter and T7 RNA polymerase. The T7 expression system is ideal for expressing soluble non-toxic recombinant protein in E. Coli. In the present work, the pET-28a (+) vector (Novagen) is used; this vector of 5.3 kb carry an N-terminal His Tag/thrombin/T7 Tag configuration plus an optional C-terminal His Tag sequence. In the present work, a schematic representation of the different cloning steps for constructing the expression plasmid for SMN protein using pET-28a (+) vector is given in Figure 5. The correct orientation for SMN

expression protein of the SMN-cDNA's insert in the plasmid vector (**5**) was confirmed by the digestion of (**5**) with NcoI which gave two fragments of 5,620 and 565 bp (data not shown).

VI. CONCLUSION

The present invention is a new approach to the cloning of human SMN gene based on the RT and PCR reactions. In comparison to the old way of isolating SMN-cDNA clones from the cDNA library by using synthesized oligonucleotide probes, the procedure for cloning of human SMN gene via RT –PCR reactions developed herein is cost-effective, not time-consuming, and is suited for any laboratory. As a result, the construction of the expression plasmids for human SMN protein can be performed by any laboratory. The present invention makes it easier to obtain full-length SMN protein which is valuable for biochemical and biological analyses that may elucidate the molecular mechanism of SMA. Knowing the molecular mechanism of SMA will allow the exploration of gene therapy in SMA.